SITUATIONAL JUDGMENT TEST RESPONDING: BEST AND WORST OR RATE EACH RESPONSE

A Thesis

by

JENNIFER LYNN RASMUSSEN

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

May 2009

Major Subject: Psychology

SITUATIONAL JUDGMENT TEST RESPONDING: BEST AND WORST OR RATE EACH RESPONSE

A Thesis

by

JENNIFER LYNN RASMUSSEN

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Approved by:

Chair of Committee,	Mindy E. Bergman
Committee Members,	Winfred E. Arthur, Jr.
	Murray R. Barrick
Head of Department,	Leslie C. Morey

May 2009

Major Subject: Psychology

ABSTRACT

Situational Judgment Test Responding: Best and Worst or Rate Each Response.

(May 2009)

Jennifer Lynn Rasmussen, B.S., Florida State University

Chair of Advisory Committee: Dr. Mindy E. Bergman

This study explores the differential validity of SJT responding formats (i.e., selecting a response to an SJT item). It was hypothesized that the SJT on which respondents identified the best and worst options would be related to cognitive ability because this type of SJT has a high cognitive load and thus acts like a knowledge test. It was also hypothesized that the SJT on which respondents rated the effectiveness of each option on a Likert scale would be related to personality because it taps into test taker's behavioral tendencies. Results show that the best-and-worst SJT was not related to measures of education (a proxy for cognitive ability) or measures of personality. The SJT on which respondents rated the effectiveness of each option on a Likert scale was related to measures of personality but not education. Finally, because the Likert SJT has a greater number of responses, it was rescored as following the best-and-worst SJT keying. This converted SJT was related to neither measures of education nor measures of personality. The choose best-and-worst SJT significantly predicted performance, while the Likert SJT and a Likert SJT converted to a best-and-worst SJT did not predict performance.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Bergman for her continued support and advice throughout not only this project but also the rest of my graduate studies. I would also like to thank my committee members, Dr. Arthur and Dr. Barrick for their insight and advice on this project. Finally, thanks go to Dr. Jeff Weekley and Kenexa who provided me with data for this project.

On a personal note, I would like to thank my parents, Kathy and Irv Rasmussen for their love and support.

TABLE OF CONTENTS

Page

LIST OF TABLES

1. INTRODUCTION: SITUATIONAL JUDGMENT TEST RESPONDING: BEST AND WORST OR RATE EACH RESPONSE

Situational judgment tests (SJTs) arose from the principle of behavioral consistency (Wernimont & Campbell, 1968), which states that past behavior is likely to predict future behavior. SJTs are a method by which test takers are presented with prototypical work-related problems followed by a set of possible solutions. The test taker then is asked to judge the effectiveness of the proposed solutions. These responses are often used to predict performance on the job (Weekley & Ployhart, 2006).

SJTs appear to be valid in predicting performance in a variety of jobs (Chan & Schmitt, 2002; McDaniel, Morgeson, Finnegan, Campion, Braverman, 2001; McDaniel & Nguyen, 2001; Motowidlo, Dunnette, & Carter, 1990; Motowidlo & Tippins, 1993; Weekley & Jones, 1999). Also, SJTs show incremental validity in the prediction of job performance beyond measures of cognitive ability, job experience, and Big Five personality (Chan & Schmitt, 2002; Clevenger, Pereira, Weichmann, Schmitt, & Harvey, 2001; Weekley & Jones, 1999; Weekley & Jones, 1997). Although any test can show adverse impact when in a particular use, there is a general consensus that SJTs exhibit lower subgroup differences than other, traditional measures such as paper-and-pencil cognitive ability tests (Chan & Schmitt, 1997; Clevenger et al., 2001; McDaniel & Nguyen, 2001; Motowidlo et al., 1990; Motowidlo & Tippins, 1993; Weekley & Jones, 1999; Weekley & Jones, 1997). This previous research is presented as a review of the existing SJT literature. It is important to note that these studies make comparisons between methods and constructs, which confounds the interpretation of the results (Arthur & Villado,

---

This thesis follows the style of *Journal of Applied Psychology*.

2008; Campbell & Fiske, 1959). The distinction between constructs and methods is explored further in section 3.

2. SJT CHARACTERISTICS

Before discussing relevant research on SJTs, several terms need to be defined. A *test item* is defined as a single situational judgment scenario consisting of a job-related situation and several options or possible courses of action. A *test* is a collection of items used as an assessment to derive a single score. A *response* is the option that the test taker chooses. Likewise, *responding* refers to selecting an option to an SJT item, whereas *responding formats* are information in the testing environment that directs test takers to respond to the test in a particular way. A *key* is the pattern of options defined as correct and incorrect by test administrators that is used to derive the test taker's score. Finally, *format* refers to the way in which SJT items are presented. For instance, the format of an SJT could be paper-and-pencil or video-based.

SJTs are typically developed by asking subject matter experts (SMEs) such as managers or job incumbents about job-related critical incidents (Weekley, Ployhart, & Holtz, 2006). The critical incidents that are part of the performance domain of interest, such as problem-solving or leadership, are developed into test items. Then, SMEs are asked to provide examples of superior and inferior responses to the test questions. These responses are then compiled to create the responses to the hypothetical situations (Olson-Buchanan, Drasgow, Moberg, Mead, Keenan, & Donovan, 1998). A key is then constructed, either a priori or post hoc, to score test-takers' responses (Bergman, Donovan, Drasgow, Henning, & Juraska, 2006). Types of keys that are used with SJTs are discussed later in this paper.

Although an SJT can be constructed in a relatively straightforward way, there is evidence that response formats can affect test scores, even when the content of the test is well developed and held constant across testing situations (Ployhart & Ehrhart, 2003). Although there are many ways to have test takers respond to the items, two formats are very popular: choosing the best

response out of the list of presented solutions, and choosing the best and the worst options, both earning points if correctly identified (Weekley & Jones, 1999). Other forms include choosing the worst option, only rating all options for an item, or rating agreement with a single presented solution. Many advantages of using the "pick best and worst" over the single best or worst choice have been identified. Weekley et al. (2006) argued that picking the best and worst options increases the variance in test scores. The increase in variance could then have an effect on the validity of the SJT. There is also a method by which the test taker is given the hypothetical situation with a solution and then asked to rate the effectiveness of the solution on a Likert-type scale (Chan & Schmitt, 2002; Ployhart & Ehrhart, 2003). Tests with this form of responding have been found to have incremental validity in the prediction of performance over measures of cognitive ability, the Big 5 personality traits, and job experience (Chan & Schmitt, 2002), as well as increased predictive validity due to the increased variance associated with the number of scorable items (Weekley et al., 2006). This is due to the fact that with the Likert-type format, each response option essentially becomes its own item, increasing the number of items in the SJT. SJTs which are comprised of one item with several responses are the most commonly used type of SJT, and are the focus of the current study.

This study focuses on SJT responding formats. Specifically, this study will examine whether different types of responding formats affect the validity of SJTs. Recent studies have shown that differences in other SJT characteristics, such as scoring keys and types of response instructions, affect the construct-related validity of SJTs (Bergman et al., 2006; Kanning, Grewe, Hollenburg, & Hadouch, 2006; McDaniel, Whetzel, Hartman, Nguyen, & Grubb, 2006; Weekley, Ployhart, & Holtz, 2006). Differences in the components of SJTs can fundamentally change what the SJT is measuring because these components affect the responses elicited from

test takers. Following this line of reasoning, it is reasonable to expect that test taker responding format will also have an effect on validity. To explore this idea, these other design factors and their influence on the validity of SJTs will be reviewed. Also, the construct/method distinction in SJT research is examined, something that has been largely overlooked by SJT researchers. Then, this study will explore the possible effect of response format on SJT validity. Finally, this paper will empirically examine these hypothesized effects.

This study will add to the current understanding of SJTs in several ways. First, this study will compare responding formats from two content invariant SJTs, something that is often confounded in other SJT studies. Also, this study will examine the effect of different responding formats on the validity of SJTs, an area not well explored by SJT researchers. This study also furthers our understanding of SJT scoring, as an SJT will be scored in multiple ways. Thus, this study explores how SJT administration decisions affect the validity of SJT assessments.

3. WHAT DO SJTS MEASURE?

SJTs seem to be valid in predicting a number of work-related outcomes, but what exactly do they measure? First, it is important to distinguish between SJTs as a method and as a construct in order to understand what SJTs really measure and to make sense of the validity findings presented later in the paper. A construct refers to the behavioral domain which is being measured. A method refers to the means by which the construct is evaluated. Making comparisons between methods and constructs confounds the interpretations of results (Arthur & Villado, 2008). While the predictor construct/method differentiation is important, the distinction in the literature has not always been made. Methods should be compared with other methods and constructs should be compared with constructs. In general, predictor assessments can be conceptualized as both methods and constructs. Sometimes, however, a particular predictor method can have an effect on the construct being measured. This is termed construct-irrelevant variance (Messick, 1995). In the case of SJTs, a "judgment" construct is probably being measured along with the intended work-related constructs. This "judgment" construct will be discussed in detail later in the paper. SJT administrators who only consider SJTs as a method of measurement are probably ignoring this "judgment" construct resulting in construct-irrelevant variance.

It is necessary to conduct studies so that SJTs as a method are compared to other methods (and SJT methods are compared to other SJT methods), holding the constructs constant, in order to properly determine the variance components (Campbell & Fiske, 1959). If constructs are compared to methods when determining effects on validity, then it is impossible to determine whether the effects are due to the method or the construct. It is important to note that it is difficult to implement such studies, and very few authors have correctly done so. Two papers

have investigated SJT methods by comparing video-based and paper-and-pencil SJTs, while keeping the constructs constant, to determine subgroup differences and criterion-related validity (Chan & Schmitt, 1997; Lievens & Sackett, 2006). These studies showed that video-based SJTs exhibit lower subgroup differences and higher predictive validity than paper-and-pencil SJTs. The method/construct distinction is one that is relatively new to the domain of SJTs, so researchers should be careful when designing studies to only compare either methods or constructs so that evidence for convergent and discriminant validity can be properly interpreted (Arthur & Villado, 2008).

There is some evidence that suggests that SJTs can be conceptualized as a method of measurement. SJTs can be designed to assess a number of different constructs including interpersonal skills, job knowledge, and leadership (Ployhart & Ehrhart, 2003; Chan & Schmitt, 2002; Clevenger et al., 2001). Critical incidents are compiled from SMEs relating to specific knowledge, skills, and abilities (KSAs) needed on the job. These critical incidents are then transformed into scenario-based questions and a list of possible responses. This results in SJTs being considered a method of measurement much like a multiple-choice test or a video-based test. Therefore, SJTs can be used as a low fidelity method to gain information about prospective employees' knowledge and application of knowledge on the job (Gessner & Klimoski, 2006). Evidence to support the assertion that SJTs are a method is seen in the claim that SJTs are job-specific (Gessner & Klimoski, 2006). In other words, SJTs are designed in order to reflect the necessary KSAs for a particular job. Because each SJT developed measures different critical incidents for each job, it can be assumed that SJTs are a method of measurement which can reflect various KSAs, which are the constructs.

However, it has also been asserted that by its very nature, an SJT is also assessing its own unique construct in addition to intended KSAs. Chan (2006; Chan & Schmitt, 2006) argued that by presenting individuals with job-relevant situations and asking them to choose the best course of action, SJTs may be assessing test taker judgment along with the KSAs being assessed. This "judgment" factor is expected to be related to cognitive ability (due to the fact that judgment has an analytical component), but also related to creativity, previous experience with the subject matter, and an ability to predict future outcomes of behavior. Therefore, although "judgment" is expected to be related to cognitive ability, it is not the same as cognitive ability because it also encompasses other constructs. In order to successfully answer an SJT item, test takers must possess the characteristics being assessed by the item, but they also have to be able to answer within the context of a particular situation, and therefore need to possess some sort of good judgment trait (Chan, 2006; Chan & Schmitt, 2006). Because "judgment" is part of the variance, it is expected that it would be related to scores on most SJTs. Therefore, it is asserted that SJTs are a method of measurement, but due to their methodology, they also contain a judgment construct that may account for some of the variance in test scores. For example, although an SJT might be designed to assess interpersonal skills, the construct-irrelevant variance may also capture another construct. However, there is lack of construct-related evidence for SJTs as a whole, which makes it unclear as to what comprises the construct-irrelevant variance. The absence of construct-related validity evidence is not something new. SJTs, like other personnel selection tools such as assessment centers and interviews, exhibit content- and criterion-related validity, but seem to supposedly display questionable construct-related validity evidence, known as a validity paradox (Arthur, Day, & Woehr, 2008; Arthur, Woehr, & Maldegen, 2000; Woehr & Arthur, 2003).

Several theories have been suggested as to the nature of dominant SJT constructs. First, Gessner and Klimoski (2006) hypothesized that SJTs may tap into several constructs, specifically: reasoning ability, creativity, knowledge of procedures, skills on the job, and ability to "guess" what answer the test designer thinks is correct. So while SJTs may be measuring intended constructs (KSAs), they may also be measuring a general "reasoning" construct due to the fact that test takers need to compare the relative impact of various options. Also, because SJTs are designed to measure how a person acts in a situation where there are many plausible answers, SJTs may measure creativity in how to best handle these unique situations. Therefore, one would expect that SJTs would exhibit some relationship with general mental ability (GMA), regardless of their specific content. This is evidenced by McDaniel et al. (2001) who reported a meta-analytic correlation of .46 between various SJTs and cognitive ability. This correlation is based on SJTs that have a high degree of variability around the mean. So while SJTs probably show some relationship with GMA, it may be that the actual strength of the relationship varies depending upon the particular construct being measured. One would expect that the relationship between an SJT and GMA would be higher when measuring constructs that are highly correlated with GMA, such as facets of knowledge (Hunter, 1986).

Brooks and Highhouse (2006) proposed that SJTs measure individual differences in judgment. Focusing on the judgment and decision-making literature, the authors identified several criteria that make up good judgment. They argued that good judgment is based equally in intuition and analysis of the situation. SJTs function by asking test-takers to predict what they would or should do in a given situation or to predict the consequences of each test option. This requires that a test taker have good judgment and decision-making skills in order to correctly answer the test item. Factors such as test taker mood, overconfidence in answers, unrealistic

optimism, and situational construal are also expected to affect making accurate judgments (Brooks & Highhouse, 2006). Understanding these and other judgment and decision-making processes will help researchers to be able to construct better SJTs.

Further, Ployhart (2006) emphasized the multidimensionality of SJTs, meaning that like other predictors such as assessment centers and interviews, SJTs can measure many latent constructs in a single assessment. Ployhart suggested that whereas researchers can focus on correlations between SJTs and other predictors, additional evidence of validity can be gained by focusing on response processes. One important contribution of Ployhart's theory is the emphasis on examining SJT for forms of construct-irrelevant variance. This is the foundation toward identifying the constructs being measured by SJTs. Ployhart argues that latent constructs (such as the test-intended KSAs) are measured by SJTs, but there are also sources of construct-irrelevant test variance as well as influences of test-taker motivation that factor into test outcomes.

Alternatively, Stemler and Sternberg (2006) authors asserted that SJTs measure an aspect of practical intelligence. Stemming from Sternberg's work on the triarchic theory of intelligence (Sternberg, 1985), practical intelligence is defined as the "ability to adapt to, shape, and select real-world environments" (Stemler & Sternberg, 2006, p. 109). Practical intelligence involves the construct of knowledge, both explicit and tacit. Stemler and Sternberg (2006) argue that SJTs measure tacit knowledge, the kind of knowledge that guides people's behavior but which people are often unable to state explicitly. Tacit knowledge influences the test taker in three areas of SJT responding: the situation underlying the problem to be faced, response strategies, and the cultural context. For example, tacit knowledge may influence how the test taker perceives the situation (e.g., whether the situation calls for insubordination or an exertion of status), how the test taker responds (e.g., whether the test taker should choose to comply with another's requests or

delegate tasks), and finally how to interpret the culture of the situation (e.g., is the situation within an organizational context, or an educational context). Similarly, Chan and Schmitt (2006) believe that SJTs may be tapping into a form of practical intelligence.

I believe that SJTs are a form of assessment that, while conceptualized as a method of the measurement of KSAs, also assess some form of judgment or decision-making construct. SJTs can be developed to assess various job-related constructs including past job experiences and job knowledge. However, due to the very nature of assessment, a situation with many plausible answers, SJTs are also testing some sort of ability to anticipate future outcomes of behavior, a component of judgment. They involve test takers' ability to infer events beyond what is written in the test item. Test takers must interpret the situation given and the consequences of each of the response options in order to correctly answer the question. This involves a judgment ability that may not be intended in the original assessment. Considering SJTs only as a method of measurement ignores the construct-irrelevant variance related to the "judgment" construct. In other words, if SJTs are used to assess intended constructs (such as various KSAs) it is important to note that they are also simultaneously measuring a construct related to some sort of adaptable, practical knowledge (Chan & Schmitt, 2006). SJT researchers need to be cognizant of the construct-irrelevant variance so that it can be eliminated or even used if it is predictive.

Historically, the literature on SJTs has almost exclusively focused on content and criterion-related validity. Research concerning the construct-related validity of SJTs has been sparse at best, so it is accurate to say that we really do not fully understand what SJTs measure. Several theories have been presented but there is a general lack of data to support any of these theories. Research has shown that SJTs have criterion-related and content-related validity in predicting job performance, which is paramount for users of SJTs. Researchers need to

determine the construct-related validity in order to better understand SJTs. Theories concerning the construct-related validity of SJTs needs to be understood so that more comprehensive and psychometrically sound SJTs can be developed. We also need to be fully aware of what is being measured in order to understand the role of subgroup differences and adverse impact in SJTs and to be able to justify their use in personnel selection.

4. TEST-BASED FACTORS AFFECTING SJT VALIDITY

4.1 Scoring

Because SJTs do not have a "true" correct answer, one of the main questions in SJT research is which scoring key to use. Different ways of scoring SJTs could have an effect on the validity of an SJT. From a review of six different scoring strategies (empirical, theoretical-rational, expert, hybridizing, factorial, and subgrouping) found in the SJT and biodata literatures, Bergman et al. (2006) developed eleven different keys for a single set of responses to one multiple choice SJT. That is, the items, options, content, response instructions, and responses were kept the same; only the manner in which correct and incorrect options were identified differed. Their work showed a wide range of validity coefficients across keys with the criterion of supervisory performance ratings. The various keys also had different discriminant validities with measures of cognitive ability and Big Five personality. Further, incremental validity in criterion prediction, above and beyond cognitive ability and Big Five, varied across keys, although it was generally higher for those keys that had higher validity coefficients. Together, these results showed that by simply applying a different scoring method to the same test with the same responses from the same people, validity varied widely. So, researchers may erroneously conclude that SJTs are not valid, when it is the key that is not valid.

4.2 Format

The format through which applicants take an SJT can affect the way the test is interpreted. Kanning et al. (2006) investigated face validity of SJTs by measuring test takers' reactions to SJTs. They found that high fidelity SJTs, which include interactive and video-based SJTs, were perceived to be more realistic and useful than text-based formats. McDaniel et al. (2006) found that high fidelity SJTs have smaller subgroup differences than low fidelity or

written SJTs, presumably because of the decrease in cognitive demands associated with the reading requirements for SJTs. Finally, Lievens and Sackett (2006) found that video-based SJTs had higher predictive and incremental validity than written SJTs when testing interpersonal and communication skills.

4.3 Response Instructions

Response instructions also can influence the validity of SJTs (McDaniel, Hartman, Whetzel & Grubb III, 2007; Ployhart & Ehrhart, 2003; Weekley et al., 2006). Ployhart and Ehrhart (2003) identified several different response instructions from the SJT literature and speculated that each might measure a different construct. In some instances, test takers are asked to choose the best response whereas in others they are asked to choose the response signifying which of the options they would be most likely to engage in for the given situation. The former asks the test taker to identify what they "should do" whereas the latter asks test takers to choose what they "would do" in the given situation (Ployhart & Ehrhart, 2003). These are often referred to as knowledge and behavioral tendency instructions, respectively (McDaniel et al., 2007). McDaniel et al. (2007) argued that because of the way the instructions are worded, behavioral tendency questions should measure typical or average performance, whereas knowledge instructions measure maximal performance. They suggested that behavioral tendency test scores would be related to personality measures because personality measures are self-reports of behavioral tendencies. That is, behavioral tendency SJTs and personality measures both ask individuals what they would do either in general (personality measures) or in a specific situation (SJTs). In contrast, McDaniel et al. (2007) argued that knowledge is one facet of cognitive ability; therefore, SJTs administered with knowledge instructions should be related to cognitive ability.

There is some evidence to support these views. McDaniel et al. (2007) found that SJTs with behavioral tendency instructions were more highly correlated with conscientiousness, agreeableness, and emotional stability (.30, .33, and .31, respectively) than were SJTs with knowledge instructions (.21, .17, .10, respectively). SJTs with knowledge instructions had higher correlations with cognitive ability (.32) than did tests with behavioral tendency instructions (.17). Ployhart and Erhart (2003) found that knowledge instruction SJTs had higher means and were more negatively skewed than behavioral tendency instruction SJTs when measuring academic study skills; knowledge instruction responses also had better criterion-related and construct validity. Further, Ployhart and Erhart showed that scores for one SJT repeated under the two response instructions (i.e., taken twice by the same respondents) were weakly correlated, even though the test content was kept constant. McDaniel et al. (2001) found that SJTs are related to cognitive ability, agreeableness, conscientiousness, and emotional stability to differing degrees based on response instructions. These studies demonstrate that the wording of response instructions can change the meaning of the test.

Nguyen, Biderman, and McDaniel (2005) studied the effects of different response instructions on the fakability of SJTs. Participants were asked to respond to both behavioral tendency and knowledge instructions by either honestly responding or faking good. Nguyen et al. (2005) found that knowledge tests had higher correlations with cognitive ability and were more difficult to fake than were behavioral tendency tests. This may be further evidence of construct differences across SJTs with different response instructions, especially the contention that different instructions tune test-takers to knowledge or to behavioral tendencies.

5. THE CURRENT STUDY

The purpose of the current study is to examine how the criterion-related validity of an SJT is affected by responding format. In this study, one multiple choice SJT was used with two different responding formats: (1) choose best-and-worst and (2) rate each option on a Likert scale. The content, the items, and their five options were the same across responding format administrations; only the responding format varied.

It is possible that the responding format affects SJTs much in the same way as response instructions. Some formats could encourage knowledge-based responding whereas others could encourage behavioral tendency responding because they focus the test taker on different cognitions. For example, in the choose best-and-worst format, test takers must evaluate the relative value of each option for an item. Comparisons are made across options to identify which is best and which is worst, producing evaluations of the most appropriate and least appropriate way to act. Multiple comparisons such as these may create test items which have a high cognitive load for the test taker. This high cognitive load is related to the construct-irrelevant variance mentioned in section 3. Therefore, those who have higher cognitive ability may be able to better answer these high cognitive load test items, similar to findings from education research which suggest that ability determines the amount of cognitive load a person is able to handle (Kashihara, Hirashima, & Toyoda, 1994). This higher cognitive load is likely related to, but not exactly the same as the "judgment" factor as judgment is thought to be comprised of other things besides cognitive ability. Also, the choose best and worst SJT asks the test taker to choose the maximal action for every test item, either maximally good performance as in the case of choose the best item or maximally bad as in choose the worst option. Thus, the choose best and worst format might be a knowledge test.

While choose best and worst SJTs force the test taker to choose one item at a time (first choose the best option, then choose the worst option), Likert SJTs present a situation in which each response option is evaluated on its own merits. This may elicit thoughts about the efficacy of each option from past experiences. Similar to McDaniel et al.'s (2007) behavioral tendency SJTs, Likert SJTs may measure typical performance because test takers must think of how they actually handled similar situations in the past. If Likert SJTs address behavioral tendencies, then they should be related to personality, another behavioral tendency. In other words, people with similar personalities should have chosen similar actions in the past, so they may be inclined to respond similarly to SJT items. Further, options do not have to be compared to produce a rating of any single option; each can be evaluated on its own merit. Even though the "maximally best" and "maximally worst" options can be identified post-responding based on the given ratings, the test-taker does not have to determine these while responding to the test items. In the case of Likert SJTs, cognitive loads (part of the "judgment" construct) might not be as high as for best and worst SJTs, and typical as opposed to maximal performance might be measured. Therefore, it is predicted that Likert SJTs elicit a behavioral tendency test.

Evidence to support these contentions would mirror previous findings regarding knowledge and behavioral tendency SJTs. Previous research demonstrated that knowledge tests are more strongly related to cognitive ability and to job performance than are behavioral tendency tests, yet behavioral tendency tests are more strongly related to personality than are knowledge tests (McDaniel et al., 2007; Ployhart & Ehrhart, 2003). In order to test this, the test content will be invariant, so that the constructs intended to be measured are the same. Therefore,

Hypothesis 1: The choose best-and-worst format is positively related to a) performance, b) cognitive ability, and c) personality.

Hypothesis 2: The Likert rating format is positively related to a) performance, b) cognitive ability, and c) personality.

Hypothesis 3: The choose best-and-worst format is more strongly related to a) cognitive ability and b) performance than the Likert format.

Hypothesis 4: The Likert format is more strongly related to personality than is the choose best-and-worst format.

It is important to note that these formats likely also differ in their variability due to the number of responses made and the ways that they are made. Therefore, a control on variability may be useful for a stronger test of these propositions. To do so, it is possible to convert one of the test formats (Likert-rating) to an analogous version of the other (choose best-and-worst). Simply, it is possible to identify the highest and lowest Likert-rated options for an item and to apply the key from the best-and-worst scoring to these options as though they were chosen in the best-and-worst format. This approach preserves the effect of the cognitions that the Likert-rating format elicited in test takers, yet also equalizes the effect of scoring and keys across the two formats. Although the strength of the relationships between Likert format and cognitive ability, personality, and performance criterion will be attenuated when the Likert responses are converted to the best-and-worst key, the overall pattern of results predicted in Hypotheses 2, 3, and 4 are still expected. These are Hypotheses 5, 6, and 7.

Hypothesis 5: The converted Likert rating format is positively related to a) performance, b) cognitive ability, and c) personality.

Hypothesis 6: The choose best-and-worst test is more strongly correlated with a) cognitive ability and b) performance than the converted Likert format.

Hypothesis 7: The converted Likert format is more strongly correlated with personality than is the choose best-and-worst format.

The current study uses an SJT that is comprised of several domains. Therefore, it is necessary to not only analyze the overall predictive validity of the SJT, but also the validity of the leadership, interpersonal skills, and sales knowledge subscales of the SJT. Each of the subscales should predict performance as they reflect the knowledge and skills that are required to succeed on the job (Jeanneret, Borman, Kubisiak, & Hanson, 1999). However, sales knowledge is most likely to affect sales performance, followed by interpersonal skills. Unless test takers of the SJT are also supervisors, good leadership skills are probably the least strong predictor of sales performance. Hunter (1986) reported a meta-analytic correlation of .64 between job knowledge and cognitive ability, therefore sales knowledge and cognitive ability are expected to be related in this study.

With regard to personality, research has shown that leadership is related to conscientiousness, extraversion, openness to experience, and neuroticism (Judge, Bono, Ilies, and Gerhardt, 2002). Based on Judge et al. (2002), the most important determinant of leadership effectiveness based on meta-analytic evidence was extraversion (.31), followed by conscientiousness (.28), neuroticism (-.24), and openness (.24). Agreeableness had the smallest correlation with leadership (.08). Judge et al. (2002) also analyzed some lower-order personality factors that are relevant to this study. Sociability (.37) was highly related to leadership as well as adaptability (.35) and dependability (.30). With respect to the O*Net personality traits used in this study, it is therefore expected that the leadership subscale of the SJT will be related to personality, particularly in the areas of social orientation, adaptability, analytical orientation,

dependability, detail orientation, persistence, and self-control. Additionally, the leadership subscale should be highly related to the leadership personality trait.

With respect to the interpersonal skills subscale, Alge, Gresham, Heneman, Fox, and McMasters (2002) found a strong relationship between interpersonal skills and extraversion (.65). Additionally, Muchinsky (1993) found relationships between Hogan Personality Inventory traits and interpersonal skills. The highest predictor of interpersonal skills was adjustment (.38), managerial potential (.30), clerical aptitude (.29), service orientation (.25), and likability (.17). These traits correspond roughly to the O*Net compentencies of adaptability, analytical, dependability, detail orientation, leadership, and social orientation.

Less is known about the relationships between personality and job knowledge. Motowidlo, Brownlee, and Schmit (2008) found that conscientiousness (.14), agreeableness (.18), and extraversion (.24) were related to customer service knowledge. Although it seems that personality may be related to job knowledge, the personality traits related to job knowledge are probably content specific. This study concerns sales knowledge, something comprised of various procedures about communication/negotiation, preparing reports, and identifying good customers. With respect to the O*Net competencies, sales knowledge should be related to achievement, detail orientation, persistence, self-control, and social orientation. Therefore,

Hypothesis 8: The leadership, interpersonal skills, and sales knowledge subscales of the SJT are positively related to a) sales performance, b) cognitive ability, and c) personality.

Hypothesis 9: Magnitudes of validity coefficients in predicting sales performance vary across subscales, with the strongest effect from sales knowledge, followed by interpersonal skills, and then leadership.

Hypothesis 10: The sales knowledge subscale is more related to cognitive ability than the leadership and interpersonal skills subscales.

Hypothesis 11: The a) leadership, b) interpersonal skills, and c) sales knowledge subscales will be related to overall personality overall. The overall relationships are expected to be predicated on d) leadership being related to leadership, analytical, social orientation, adaptability, dependability, detail orientation, persistence, and self-control, e) interpersonal skills being related to adaptability, analytical, dependability, detail orientation, leadership, and social orientation, and f) sales knowledge being related to achievement, detail orientation, persistence, self-control, and social orientation.

Although the relative validities of SJTs under various responding and scoring formats are important, it is also important to examine the incremental validities of these tests above and beyond cognitive ability and personality. Cognitive ability tests are among the most prominent selection test devices because they have a long and strong history of prediction of job performance (Hunter, 1986; Hunter & Hunter, 1984; Schmidt & Hunter, 2004). Personality tests are also often used in selection batteries and have been shown to predict job performance (Barrick & Mount, 1991; Hough, 1992; Salgado, 1997). Many cognitive ability and personality assessments are commercially available and are relatively inexpensive to use in selection procedures. SJTs, on the other hand, are less common and probably more costly to develop. Thus, it is important to show that SJTs contribute to the prediction of job performance beyond cognitive ability and personality tests. These hypotheses relate to the incremental validity of the SJT and not just the relationship between the variables.

Hypothesis 12: The choose best-and-worst format has incremental validity in the prediction of job performance above and beyond cognitive ability and personality.

Hypothesis 13: The Likert format has incremental validity in the prediction of job performance above and beyond cognitive ability and personality.

Hypothesis 14: The converted Likert format has incremental validity in the prediction of job performance above and beyond cognitive ability and personality.

Hypothesis 15: Each of the subscales of the SJT (leadership, interpersonal skills, and sales knowledge) has incremental validity in the prediction of job performance above and beyond cognitive ability and personality.

6. METHOD

6.1 Participants

Salespeople working for two U.S. business-to-business sales companies participated in this study. Complete, usable data from Company 1 were gathered from 169 current employees who had at least six months of experience on the job. The sample was 93.5% Caucasian and 6.5% ethnic minorities. Additionally, 44.8% were female and 33.7% were over 40. Job tenure averaged 4.60 years (SD= 0.87).

Complete, usable data from Company 2 were gathered from 184 employees who had at least twelve months of experience on the job. The ethnic composition of the sample was 81.0% White, 14.4% African-American, 2.9% Hispanic, and 1.7% Asian. Additionally, 47.7% were female and 52.3% were male. Average job tenure for Company 2 (M= 3.92 years, SD= 1.17). On average, employees from Company 2 were 39.21 years of age (SD= 9.38).

To ensure that there was adequate statistical power, a power analysis was conducted post-hoc using the program G*Power (Faul, Erdfelder, Lang, & Buchner, 2007). The power analysis was done post-hoc based on meta-analytic correlations. Medium effect sizes were expected, based on the size of the effect sizes anticipated based on meta-analytic correlations. McDaniel et al. (2001) found that the correlation between SJTs and job performance was .36. Statistical power for the overall best-and-worst SJT was 0.57 and statistical power for the overall Likert SJT and converted Likert SJT with performance was .61. These power levels are below the .80 convention, meaning that there is a relatively high level of risk of committing Type II error (Cohen, 1992).

In both companies, all participants completed the SJT on company time via the Internet as part of a larger predictor battery. Company 1's battery took approximately 120 minutes to

complete, whereas Company 2's took approximately 90 minutes to complete. For each company, criterion information was collected from organizational records.

6.2 Measures

6.2.1 Situational Judgment Test

The SJT was comprised of 52 items designed to tap into such constructs as sales knowledge, leadership, and interpersonal skills. The sales knowledge subset consisted 25 items, the leadership subset consisted of 9 items, and the interpersonal skills consisted of 18 items. Each item was followed by five options. For example, one sales knowledge item asked test-takers to choose the best response for a situation in which a client—at the final contract negotiation—says that a deal could be off if more services are not added to the contract at no additional cost, even though you've reiterated that the fee in the contract is the lowest allowed by policy. The response options are: (1) review the benefits of the existing services to dissuade the client from asking for new services; (2) stick to the existing contract while explaining that you cannot break corporate policy; (3) get your supervisors involved; (4) change the corporate policy; and, (5) negotiate a small increase in the fee so that the desired services can be delivered. For this SJT item, the correct best answer is (5) negotiate a small increase in the fee so that the desired services can be delivered and the correct worst answer is (4) change the corporate policy.

In Company 1, test takers chose both the best and the worst response for each item. Therefore, a total of 104 responses were made by each Company 1 test taker (two per item). In Company 2, test takers were instructed to rate each option on a five point scale (1= Most Ineffective to 5= Most Effective); thus, a total of 210 responses were made by each Company 2 test taker (one per response option, five response options per item).

Ten subject matter experts also completed the SJT and their mean effectiveness ratings were used to develop the scoring keys. For the best-and-worst key, the most and least effective options were identified from the SME ratings. Test-takers earned one point each for identifying the best and the worst answers; one point was lost for choosing the worst (best) answer as best (worst) (Motowidlo et al., 1990; Weekley & Jones, 1999). Thus scores on each item could range from 2 to -2 and were calculated as follows: (1) to receive a 2, a person would have to chose as their best response the "correct best answer" and as their worst response the "correct worst answer," i.e., get both right; (2) to receive a -2, a person would have to chose the "correct worst answer" as their best response and the "correct best answer" as their worst response, i.e., get both wrong; (3) a 1 was received if the respondent successfully identified either the "correct best answer" or the "correct worst answer," but not both; (4) a -1 was indicated if the respondent chose as the best response the "correct worst answer" or chose as the worst response the "correct best answer," but not both; and (5) a 0 was received by those choosing neither the correct best nor the correct worse response. These scores were derived for Company 1 participants.

For the Likert score, the absolute value was taken of the difference between the test taker's response to each option and the mean SME rating for that option. The difference values for each option were summed across all items to obtain the total score for the Likert key. This procedure produces scores that are *lower* when closer to the correct rating; therefore, these scores would be negatively correlated with the criterion measures. In order to make the results more interpretable, the directionality of the difference scores were then reversed such that higher scores indicate better scores and positive correlations with the criterion are predicted. These scores were derived for Company 2 participants.

To create the converted Likert score, the best-and-worst key was used. If the test taker gave the correct best option the highest Likert rating among the options for that item (whether or not tied with another option), it was counted as correct and the test-taker received a point. If the correct worst option received the lowest Likert rating on that item from the test-taker (regardless of ties with other options), it was counted as correct and the test-taker received a point. Points were also deducted if the keyed best and worst options were incorrectly rated as the least effective and most effective, respectively. Thus, the scores for each item could range from –2 to +2, as they did for the best-and-worst SJT scoring key in Company 1.

In order to develop the three subscales of the SJTs, the scores corresponding to the test items relating to each of leadership, interpersonal skills, and sales knowledge were calculated for each of the types of SJT responding formats in the same manner mentioned above.

6.2.2 Biodata

A biodata questionnaire was used to assess tenure and education level. Here biodata is interpreted as a method of obtaining information about prior life and work experiences, similar to the way one may use an interview to gather the same information. Responses were made on five or six point scales varying by item. For this study, a single item asking about the amount of time in the current job was used as job tenure. Questions about education (9 items) were combined to form a total education score which will be used in lieu of direct measures of cognitive ability, which were not available. These questions ask about the level of education obtained and grades in educational courses.

6.2.3 O*Net Competencies

O*Net competencies were assessed with 170 items. Consistent with the foundational O*Net research (Peterson, Mumford, Borman, Jeanneret, & Fleishman, 1999; Peterson,

Mumford, Borman, Jeanneret, Fleishman, & Levin et al., 2001), 17 competencies related to the job were assessed with ten items per scale: achievement/effort, adaptability, analytical, concern for others, integrity, detail orientation, energy, initiative, innovation, persistence, stress tolerance, self-control, cooperation, dependability, independence, leadership, and social orientation. Responses were made on a five-point Likert scale. O*Net competencies include individual worker characteristics which are relatively stable over time (Peterson et al., 2001), much like personality traits. Therefore, these O*Net competencies will be used as personality variables for this study.

6.2.4 Criterion Measures

Sales data were collected for each employee in order to measure job performance. The criterion used for both Company 1 and Company 2 was achievement to sales quota.

7. RESULTS

Means, standard deviations, correlations, and reliabilities are found in the appendix in Tables 1 (Company 1, best and worst SJT) and 2 (Company 2, Likert SJT). All of the hypotheses were tested with regressions; sex and tenure were controlled for in all of the following analyses. Because measures of cognitive ability were not available, level of education was used. Past research has shown that education is an acceptable, though not optimal, proxy for cognitive ability (Berry, Gruys, & Sackett, 2006).

Table 3 of the appendix contains the regressions of performance criteria onto each of the three SJT scores, after controlling for sex and tenure. Consistent with Hypothesis 1a, the best-and-worst SJT significantly predicted performance. However, neither the Likert SJT nor the converted Likert SJT was related to the performance criterion (Hypotheses 2a and 5a, respectively). Table 3 of the appendix also contains the regressions of the three subtests (leadership, interpersonal skills, and sales knowledge) on performance. Contrary to H8a and H9, none of the SJT subsets showed a relationship with performance for any of the best and worst SJT, the Likert SJT, or the converted Likert SJT.

Table 4 of the appendix contains regressions of the SJTs on education levels and personality dimensions. Education (as a proxy for cognitive ability) did not significantly predict the best-and-worst SJT, contrary to H1b. Neither the Likert nor the converted Likert SJT was related to education level (Hypotheses 2b and 5b, respectively). As for personality (Hypothesis 1c), the O*Net competencies were not related to the best-and worst SJT; although dependability and leadership did have significant beta weights in the regression, the step including the 17 competencies was not significant. Hypothesis 2c, relating personality to the Likert SJT, received some support, as the regression step adding all O*Net competencies was significant. Among the

competencies, leadership (β= 0.27) and persistence (β= -0.46) were significant. The converted Likert SJT was not related to personality overall (Hypothesis 5c), although leadership had a significant beta weight in the regression. The difference in results for Likert and converted SJTs suggest that the Likert SJT might be related to personality only because of increased variability due to the many responses rather than the responding format used. However, it might also be that at least leadership is important, but the number of parameters added to the model in this last step masks the importance of specific personality traits.

The results from Tables 3 and 4 of the appendix also speak to Hypotheses 3, 4, 6, and 7, as they describe the relative strength of relationships with performance, cognitive ability, and personality across the three SJTs. The best-and-worst SJT is more strongly related to performance than either the Likert or the converted Likert SJT, considering that neither Likert SJT was related to performance. Thus, there is some support for Hypotheses 3b and 6b. Additionally, in support of Hypothesis 4, the Likert SJT appears to have a stronger relationship with the set of personality variables, as measured by O*Net competencies, than does the best-and-worst SJT. In regard to Hypothesis 7, the converted Likert SJT is not more strongly related to personality than is the best-and-worst SJT, as neither were related to the O*Net competencies overall.

Table 5 of the appendix shows the relationship between the leadership subscale and education and personality (relating to H11a and H11d). For the leadership subscale, a significant effect was found between the Likert SJT and overall personality. However, the best and worst leadership subscale and the converted Likert leadership subscale did not have a relationship with personality. Therefore H11a is only partially supported. Table 5 of the appendix also shows the specific personality traits related to the leadership subscale. The best-and-worst leadership

subscale was related to analytical (.31), dependability (.27), detail orientation (-.26), and social orientation (-.26). The Likert leadership subscale was not related to any of the specific personality traits. The converted Likert leadership subscale was related to the leadership trait (.27). Therefore there is only partial support for H11d.

Table 6 of the appendix shows the relationship between the interpersonal skills subscale and education and personality (relating to H11b and H11e). The best and worst, the Likert, and the converted Likert for interpersonal skills was not related to education. The Likert interpersonal skills subscale is positively related to personality, but neither of the other interpersonal skills subscale are related to personality. Therefore, there is no support for H11b. With regard to specific personality traits, only the Likert subscale was related to personality. The traits that were related are achievement (-.28), leadership (-.27) and persistence (.63). Therefore there was partial support for H11e.

Table 7 of the appendix shows the results for the sales knowledge subscale (H11c and H11f), which is related to education for the best and worst SJT only and is related to personality for the Likert and converted Likert SJTs. Therefore, there is only partial support for H10. The best-and worst sales knowledge subscale was related to achievement (.31) and leadership (.24). The Likert sales knowledge subscale was related to leadership (-.30) and persistence (.42) and the converted sales knowledge subscale was related to cooperation (-.27), integrity (-.20), leadership (.39), and persistence (-.29). Therefore, there is partial support for H11c and H11f.

Hypothesis 8a asserted that the leadership, interpersonal skills, and sales knowledge subtests would be related to performance. None of the subtests were found to be related to performance under any responding instructions, therefore there is no support for H8a. Hypotheses 8b and 8c proposed that the three subtests would be related to cognitive ability and

personality. Results indicated that there was no positive relationship with cognitive ability and personality in general. Therefore, there is no support for H8b and H8c.

Hypotheses 12, 13, and 14, and 15 proposed that each of the SJTs and SJT subscales would contribute to the prediction of performance criteria above and beyond cognitive ability and personality. To test this, a series of hierarchical regressions was conducted. In Step 1, sex and tenure were controlled. In Step 2, education level was added. In Step 3, the block of 17 O*Net competencies were included. In the final step, the SJT was added. The results of the regressions are found in Table 8 of the appendix. For each SJT, the change in R^2 was not significant from Step 3 to Step 4; thus, none of the SJTs provide incremental validity in the prediction of performance criteria, above and beyond personality and education level. Therefore, hypotheses 12, 13, and 14 were not supported. In addition, none of the SJT subscales provided incremental validity in the prediction of performance over and above that of cognitive ability and personality. Therefore, H15 was also not supported.

8. CONCLUSIONS

8.1 SJTs and Performance

The validity of SJTs has been the subject of a number of studies (Chan & Schmitt, 2002; McDaniel et al., 2006). Factors such as the scoring of SJTs and response instructions have already been shown to affect the validity of SJTs (Bergman et al., 2006; McDaniel et al., 2001; Ployhart & Ehrhart, 2006). This study furthers this tradition by examining whether the responding format of SJTs also affects the validity of SJTs. The choose best-and-worst format predicted job performance better than the Likert or the converted Likert SJT. This suggests that the responding format affects the validity of the test, as the content of the items was kept constant across the various formats.

It is surprising that neither the Likert nor the converted Likert SJT were related to job performance. It may be the case that the scoring used to calculate the Likert SJT score was not appropriate. This study took the absolute value of the difference scores between the SME rating and the test taker's ratings to create the Likert SJT score. However, the directionality (positive or negative) of the difference scores from the SME rating may provide important information about the correctness of the test taker's response. For example, suppose SMEs rated a particular option as a 3 on a 5-point scale. Under the system that was used to scored the Likert SJT, test takers who chose a 1 and test takers who chose a 5 would receive the same score. However, information might be gained by examining the directionality of the scores. There may be certain instances where an option, even though rated as average by the SMEs, is considered to be more correct if a person thinks it is subaverage rather than better than average. In other words, SMEs might find it satisfactory to classify the option as an average or poor option, but definitely not a good option. The test taker would need to meet or exceed the SME score. The opposite may be

true as well. It may be better to rate an option as good if the correct SME rating is average as opposed to rating it bad. Here the test taker would not be able to exceed the SME rating.

For the converted SJT score, the ties between SME ratings and/or among test-taker's responses could contribute to suboptimal scoring. Additionally, scores were produced such that if the test-taker's highest (lowest) rated option was the keyed best (worst) option, a point was given to the test-taker. However, this scoring strategy does not account for responding where the respondent gives a low rating to the keyed best answer, yet still has that option rated highest (e.g, when four options are rated "1-most ineffective" and the correct best option is rated "2-almost most ineffective"). The conversion of Likert SJTs using different keying strategies could produce different validity coefficients for the prediction of performance by SJTs. Future research should address these issues by focusing on different ways to calculate and transform SJT scores using different scoring keys. However, Waugh and Russell (2005) conducted a study which looked at the effect of using different scoring algorithms. The authors found that the different scoring algorithms only produced very small differences in validity coefficients.

8.2 SJTs and Education

Focusing on education (as a proxy for cognitive ability) in relation to the SJTs, the choose best-and-worst SJT was not found to be more related to education than was the Likert or converted SJT. It may be the case that education did not serve as an adequate measure of cognitive ability. Although education is related to cognitive ability, they are not perfectly correlated (Berry et al., 2006). Education attainment may be influenced by motivation and economic status, whereby cognitive ability may not. Also, there may be range restriction in education because most of the values fell within a very small interval of numbers. Theory suggests that the "judgment" factor (construct-irrelevant variance) of SJTs should be related to

cognitive ability. Part of that is being able to apply reasoning in different situations. Education probably does not fully capture this reasoning ability. Therefore, future research should use a direct measure of cognitive ability to see if cognitive ability is related to SJT responding instructions to better test the theory suggested in this paper. It may also be the case that best-and-worst responding instructions do not elicit a knowledge test. Test takers may find it relatively easy to weigh the relative effectiveness of each item, producing lower cognitive load.

8.3 SJTs and Personality

As predicted, personality was found to be related to the Likert SJT, but not the choose best-and-worst SJT. Personality was also not found to be related to the converted SJT. This seems to suggest that personality may play a role when participants are given Likert responding formats as opposed to best-and-worst responding formats. However, these relationships may only exist due to the increase in variability because of the sheer number of responses in the Likert format. In fact, the conversion of the Likert SJT into the best and worst SJT led to an attenuation of the effects. It was proposed that Likert SJTs might cause people to consider past experiences because people have to think of multiple options at once. However, based on the results, test takers seem to be focusing on the item at hand.instead test takers concentrate on the problem itself resulting in a focus on the present. Additionally, test takers may not have encountered a problem like the test item before; therefore, prior experiences might have less influence on responding in such cases, as there is not a simple correspondence between the events described in the item and the test-taker's experiences. Therefore, the relationship between personality and the Likert SJT is tenuous. However, this research should be replicated using traditional measures of mainstream personality theories such as the Big Five instead of O*Net competencies to ensure that these effects hold. Relationships between the Big Five measures and

outcomes are well established and the psychometric properties of many Big Five measures are well-established (Barrick & Mount, 1991; Barrick, Mount, & Judge, 2001; Cellar, Miller, Doverspike, & Klawsky, 1996). Therefore it is important to test personality as related to the Big Five so that valid comparisons can be made with existing literature.

The leadership, interpersonal skills, and sales knowledge subscales provided some interesting findings regarding personality traits. First, the leadership SJT was related to leadership, analytical orientation, dependability, detail orientation, and social orientation as hypothesized. Detail orientation and social orientation were found to be negatively related to the leadership subscale. The detail orientation trait consisted of items regarding planning and organization. Perhaps leaders (particularly transformational leaders) are more spontaneous than detail-oriented, inspiring their subordinates to the general work mission while also allowing subordinates to determine their own workday plans. The social orientation trait is related to having close friendships. Leaders probably do not want to keep close friendships with the people they lead. Part of leadership is governing others, which can be difficult if they are close friends. The personality traits related to the leadership subscale can map onto the Big Five dimensions of extraversion (i.e., social orientation, concern for others, and cooperation) and conscientiousness (dependability, detail orientation, and initiative) and to some extent neuroticism (self-control and stress tolerance) consistent with past research on leadership (Judge et al., 2002).

The interpersonal skills subscale was related to achievement, leadership, and persistence. Persistence is a valuable trait to have when dealing with other people, particularly in a sales environment. A large part of sales is convincing other people to buy a product, something that would benefit from persistence. Achievement and leadership were negatively related with interpersonal skills, suggesting that those who have good interpersonal skills are also those who

have low achievement orientation and leadership. It may be the case that people who have good interpersonal skills are ones who prefer to focus on good relationships with others instead of achieving success. With regard to leadership, perhaps sales people are taught to follow the leadership of the customer and not to lead them.

Finally, the sales knowledge subscale was related to achievement and persistence, as anticipated, but also related to leadership, cooperation (negatively), and integrity (negatively). Because these personality traits are probably are related to the type of job performed, the results show that leadership is an important part of sales knowledge. Additionally, sales is a very competitive field, so cooperation probably hinders sales efficacy. Further, employees might able to make more sales if they bend the rules to do so; that is, they may be more effective in sales if their codes of conduct and morality are somewhat flexible and amenable to situational factors, rather than being fixed and immovable. For example, highlighting the positive aspects of products and minimizing the negative features might be somewhat unethical, but also lead to good sales outcomes.

8.4 Method/Construct Distinction

Finally, several issues need to be mentioned about the method/construct distinction. It is important to make the distinction between methods and constructs in order to be able to truly understand validity coefficients. Previous SJT studies have generally failed to separate the two, particularly when it comes to topics such as validity and subgroup differences. Future research should investigate the validity and subgroup differences of SJTs by using a multi-trait multi-method matrix (MTMM; Campbell & Fiske, 1959). By comparing methods with other methods, and constructs with other constructs, the variance components can correctly be determined. This will allow researchers resolve whether the validity or subgroup differences exhibited by SJTs

truly exist. Also, future research should focus on new ways to approach the construct validity problem of SJTs (i.e., the validity paradox). These include examining group differences, the homogeneity of items, factor structures, stability of test scores, correlation with other constructs, as well as consideration of the test content (Schmitt & Chan, 2006). Evidence of substantive validity (insight into the theoretical rationales for consistencies in scores) could also prove fruitful for determining construct-related validity (Messick, 1995). The possibility of a judgment construct within SJTs would be an exciting new direction in personnel selection. Most, if not all, occupations require that people make judgments everyday. Therefore, some sort of judgment factor is likely to be related to job performance. SJTs could potentially be able to test for any work-related construct along with a special judgment trait. If this judgment construct includes aspects of cognitive ability, creativity, past experience, and ability to predict, then multiple constructs could be tested with one assessment. This could save test administrators time and money.

Although this study does not directly test whether SJTs are indeed a method of measurement that may contain construct-irrelevant variance, the results of this study can illuminate the construct/method debate. First, SJTs can be constructed to measure particular constructs (Olson-Buchanan et al., 1998; Weekley & Jones, 1997). In this study, the overall SJT was deconstructed into several subscales (leadership, interpersonal skills, and sales knowledge). So SJTs can be conceived of as a method. Secondly, if there is a general "judgment" construct being captured by SJTs, then every SJT should be somewhat related to cognitive ability, showing positive manifold. The results of this study however do not show that all SJTs, regardless of responding format, are related to cognitive ability. However, due to the constraints of using

education as a proxy for cognitive ability, this relationship may not accurately reflect the true state of the SJT as a measure of judgment skills-cognitive ability relationship.

8.5 Limitations and Future Directions

There are several limitations to this study. As mentioned, proxies for cognitive ability (education) and personality traits (O*Net competencies) were used. While education and cognitive ability are undoubtedly related, those who score lower on cognitive ability tests may still achieve higher education levels through persistence, and vice-versa. Therefore, the relationship between cognitive ability and SJTs may be attenuated when education measures are used instead of direct measures of cognitive ability. In regard to personality, relationships between the variables could be attenuated when the O*Net competencies are used as opposed to Big Five measures. Because several of the O*Net competencies may load onto the same Big Five measure, using a Big Five measure would decrease the amount of variability due to the number of traits and make the relationships between the variables stronger.

It is important to reiterate that the data were collected from incumbents instead of applicants. For this study, the use of incumbents as opposed to applicants poses a boundary condition. This study makes no claims about the use of different SJT responding instructions in personnel selection, so applicants are not necessary for this study. One should be careful, though, in extending the findings of this study to applicants, as it would be erroneous to assume that the findings of this study would be appropriate for applicants. When applicants take SJTs, they do so under high-stakes testing situations. This may change the relationship between SJTs and other variables, particularly between SJTs and personality. Under high-stakes conditions, applicants may rely more on their behavioral tendencies than people who may have previously encountered similar situations to those described in the SJT because the latter people have more knowledge.

Also, if SJTs are at all fakeable, then we may see different relationships among the variables because applicants will be more motivated to fake than incumbents (Robie, Zickar, & Schmit, 2001). So while the use of applicants may be of importance for personnel selection studies, it does not diminish the importance of the findings of this study. However, one should be careful about generalizing the findings of this study to other groups.

Research needs to expand the findings of this study to other SJT content. If SJTs are a method which contains variance due to constructs, it is important to determine whether the effects found here are also found with the same method applied to other constructs. With a single study, it is only known that this method-construct block (SJT formats/sales job-relevant judgments) shows these effects. By finding the same effects for responding formats for SJTs in other domains (e.g., customer service, managerial duties), the effect of responding formats on validity can be better understood.

Additionally, subgroup differences in the responding formats of SJTs should be studied to ensure that the use of either best-and-worst or Likert formats does not induce greater subgroup differences. Although previous research has shown that SJTs exhibit fewer subgroup differences than other tests (Chan & Schmitt, 1997; Clevenger et al., 2001; McDaniel & Nguyen, 2001; Motowidlo et al., 1990; Motowidlo & Tippins, 1993; Weekley & Jones, 1999; Weekley & Jones, 1997), subgroup differences relating directly to responding instructions should be examined. When testing for these subgroup differences, it will be necessary to compare SJTs with identical content in order to fully test the effect of the responding instructions. This is especially important if some tests have a higher cognitive or reading demand than other SJT formats. Certainly, it seems that best-and-worst SJTs have a higher cognitive load than do Likert SJTs, because best-and-worst SJTs require comparisons of several possibilities at a time whereas Likert SJTs allow

for the consideration of a single option at a time. However, future research is needed to bear out these assumptions.

Another limitation of this study is that the reliabilities for each of the different SJTs were relatively low even compared to other reported alphas. The range of alphas for overall and subscale scores were: .10 -.50 (best-and-worst), .63 - .81 (Likert SJT), and .37 -.58 (converted Likert). In general, the reliability of SJTs, as evidenced by Cronbach's alpha, are typically low unless a very large number of items are used (Chan & Schmitt, 2006). Alphas typically range from a high end of .90 to a low end of .29 (Clevenger et al., 2001; Weekley & Jones, 1997). Meta-analytic results from McDaniel et al. (2001) show that the mean alpha of SJTs is .60. Also, the reliability of each of the three subscales was incredibly low. This might have contributed to lack of support for the hypotheses.

Low statistical power is another limitation of this study. Statistical power for the overall best-and-worst SJT was 0.57 and statistical power for the overall Likert SJT and converted Likert SJT with performance was .61. These power levels are below the .80 convention. This means that there is a risk for failing to reject a false null hypothesis (Type II error). With low power, it may mean that there were effects in the data but these effects were not detected. Future studies should test for power a priori to determine the correct number of participants needed to ensure adequate power.

It is important to note that the data were collected from two different organizations instead of coming from test takers within the same organization, which could be a confound in this study. The effects found in this study could be related to company effects instead of the relationships between the variables. Future research should look at the responses from the same people, or from a matched sample within one company, to eliminate possible company effects.

8.6 Conclusion

This study adds to the understanding of SJTs in a variety of ways. First, this study looked at an unexamined aspect of SJTs, responding instructions. It also included the transformation of the scores of one SJT under one responding format into another, a technique that has not been used. The findings from this study may allow for previous SJT data to be reconstructed into a better, more valid SJT. Finally, this study looked at three different subscales of the SJT and their relationship with other variables. Together, these additions will help to create an improved test and further solidify SJTs as a legitimate personnel selection tool.

REFERENCES

Alge, B. J., Gresham, M. T., Heneman, R. L., Fox, J., & McMasters, R. (2002). Measuring customer service orientation using a measure of interpersonal skills: A preliminary test in a public service organization. *Journal of Business and Psychology, 16*, 467-476.

Arthur, W. Jr., Day, E. A., & Woehr, D. J. (2008). Mend it, don't end it: An alternate view of assessment center construct-related validity evidence. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 1*, 105-111.

Arthur, W. Jr., Woehr, D. J., & Maldegen, R. (2000). Convergent and discrimininant validity of assessment center dimensions: A conceptual and empirical re-examination of the assessment center construct-related paradox. *Journal of Management, 26,* 813-835.

Arthur, W. Jr., & Villado, A. J. (2008). The importance of distinguishing between constructs and methods when comparing predictors in personnel selection and research and practice. *Journal of Applied Psychology, 93*, 435-442.

Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and performance: A meta-analysis. *Personnel Psychology, 44,* 1-26.

Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next. *Personality and Performance, 9,* 9-30.

Bergman, M.E., Drasgow, F., Donovan, M.A., Henning, J.B., & Juraska, S.E. (2006). Scoring situational judgment tests: Once you get the data, your troubles begin. *International Journal of Selection and Assessment, 14,* 223-235.

Berry, C. M., Gruys, M. L., & Sackett, P. R. (2006). Education attainment as a proxy for cognitive ability in selection: Effects on levels of cognitive ability and adverse impact. *Journal of Applied Psychology, 91,* 696-705.

Brooks, M. E., & Highhouse, S. (2006). Can good judgment be measured? In J. A. Weekley & R. E. Ployhart (Eds.), *Situational judgment tests: Theory, measurement, and application* (pp. 39-55). Mahwah, NJ: Lawrence Erlbaum Associates.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56,* 143-164.

Cellar, D. F., Miller, M. L., Doverspike, D. D., & Klawsky, J. D. (1996). Comparison factor structures and criterion-related validity coefficients for two measures based on the Five Factor Model, *Journal of Applied Psychology, 81*, 694-704.

Chan, D., & Schmitt, N. (1997). Video-based versus paper-and-pencil method of assessment in situational judgment tests: Subgroup differences in test performance and face validity perceptions. *Journal of Applied Psychology, 82*, 143-159.

Chan, D., & Schmitt, N. (2002). Situational judgment and job performance. *Human Performance, 15(3)*, 233-254.

Chan, D. (2006). Interactive effects of situational judgment effectiveness and proactive personality on work perceptions and work outcomes. *Journal of Applied Psychology*, *91,* 475-481.

Clevenger, J., Pereira, G. M., Wiechmann, D., Schmitt, N., & Harvey, V. S. (2001). Incremental validity of situational judgment tests. *Journal of Applied Psychology, 86*, 410-417.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175-191.

Gessner, T. L., & Klimoski, R. J. (2006). Making sense of situations. In J. A. Weekley & R. E. Ployhart (Eds.), *Situational judgment tests: Theory, measurement, and application* (pp. 13-38). Mahwah, NJ: Lawrence Erlbaum Associates.

Hough, L. M. (1992). The "Big Five" personality variables-- construct confusion: Description versus prediction. *Human Performance, 5,* 139-155.

Hunter, J. E. (1986). Cognitive ability, cognitive aptitude, job knowledge, and job performance. *Journal of Vocational Behavior, 29*, 340-362.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin, 96,* 72-98.

Jeanneret, P. R., Borman, W. C., Kubisiak, U. C., & Hanson, M. A. (1999). Generalized work activities. In N. G. Peterson, M. D. Mumford, W. C. Borman, P. R. Jeanneret, & E. A. Fleishman (Eds.), *An occupational information system for the 21st century: The development of O*NET.* (pp. 105-125). Washington, DC: American Psychological Association.

Judge, T. A., Bono, J. E., Ilies, R., & Gerhardt, M. W. (2002). Personality and leadership: A qualitative and quantitative review. *Journal of Applied Psychology, 87*, 765-780.

Kanning, U. P., Grewe, K., Hollenburg, S., & Hadouch, M. (2006). From the subject's point of view: Reactions to different types of situational judgment items. *European Journal of Psychological Assessment, 22,* 168-176.

Kashihara, A., Hirashima, T., & Toyoda, J. (1994). A cognitive load application in tutoring. *User Modeling and User-Adapted Interaction, 4*, 279-303.

Landy, F. J. (1986). Stamp collecting versus science: Validation s hypothesis testing. *American Psychologist, 41,* 1183-1192.

Lievens, F., & Sackett, P.R. (2006). Video-based versus written situational judgment tests: A comparison in terms of predictive validity. *Journal of Applied Psychology, 91,* 1181-1188.

McDaniel, M. A., Hartman, N. S., Whetzel, D. L., Lee Grubb III, W. (2007). Situational judgment tests, response instructions and validity: A meta-analysis. *Personnel Psychology*, *60*, 63-91.

McDaniel, M. A., Whetzel, D. L. Hartman, Nguyen, N. T. & Grubb W. L., III. (2006). Situational judgment tests: validity and an integrative model. In J. A. Weekley & R. E. Ployhart (Eds.), *Situational judgment tests: Theory, measurement, and application* (pp. 183-203)*,* Mahwah, NJ: Lawrence Erlbaum Associates.

McDaniel, M. A., Morgeson, F. P., Finnegan, E. B., Campion, M. A., & Braverman, E. P. (2001). Use of situational judgment tests to predict job performance: A clarification of the literature. *Journal of Applied Psychology, 86*, 60-79.

McDaniel, M. A., & Nguyen, N. T. (2001). Situational judgment tests: A review of practice and constructs assessed. *International Journal of Selection and Assessment*, *9*, 103-113.

Messick, S. J. (1995). The validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50,* 741–749.

Motowidlo, S. J., Brownlee, A. L., & Schmit, M. J. (2008). Effects of personality characteristics on knowledge, skill, and performance in servicing retail customers. *International Journal of Selection and Assessment, 16*, 272-280.

Motowidlo, S. J., Dunnette, M. D., & Carter, G. W. (1990). An alternative selection procedure: The low-fidelity simulation. *Journal of Applied Psychology, 75*, 640-647.

Motowidlo, S. J., & Tippins, N. (1993). Further studies of the low-fidelity simulation in the form of a situational inventory. *Journal of Occupational and Organizational Psychology*, *66*, 337-344.

Muchinsky, P. M. (1993). Validation of personality constructs for the selection of insurance industry employees. *Journal of Business and Psychology. Special Issue: Test Validity Yearbook: II, 7*, 475-482.

Nguyen, N. T., Biderman, M. D., & McDaniel, M. A. (2005). Effects of response instructions on faking a situational judgment test. *International Journal of Selection and Assessment, 13,* 250-260.

Olson-Buchanan, J. B., Drasgow, F., Moberg, P. J., Mead, A. D., Keenan, P. A., & Donovan, M. A. (1998). Interactive video assessment of conflict resolution skills. *Personnel Psychology, 51*, 1-24.

Peterson, N. G., Mumford, M. D., Borman, W. C., Jeanneret, P. R., & Fleishman, E. A. (Eds.). (1999). *An occupational information system for the 21st century: The development of O*NET,* Washington, DC, US: American Psychological Association.

Peterson, N. G., Mumford, M. D., Borman, W. C., Jeanneret, P. R., Fleishman, E. A., & Levin, K. Y. et al. (2001). Understanding work using the occupational information network (O*NET). *Personnel Psychology, 54*, 451-492.

Ployhart, R. E. (2006). The predictor response process model. In J. A. Weekley & R. E. Ployhart (Eds.), *Situational judgment tests: Theory, measurement, and application* (pp. 83-105). Mahwah, NJ: Lawrence Erlbaum Associates.

Ployhart, R. E., & Ehrhart, M. G. (2003). Be careful what you ask for: Effects of response instructions on the construct validity and reliability of situational judgment tests. *International Journal of Selection and Assessment*, *11*, 1-16.

Robie, C., Zickar, M. J., & Schmit, M. J. (2001). Measurement equivalence between applicant and incumbent groups: An IRT analysis of personality scales. *Human Performance, 14*, 187-207.

Salgado, J. F. (1997). The five factor model of personality and job performance in the European community. *Journal of Applied Psychology, 82,* 30-43.

Schmidt, F. L., & Hunter, J. (2004). General mental ability in the world of work: Occupational attainment and job performance. *Journal of Personality and Social Psychology, 86*, 162-173.

Schmitt, N., & Chan, D. (2006). Situational judgment tests: Method or construct? In J. A. Weekley & R. E. Ployhart (Eds.), *Situational judgment tests: Theory, measurement, and application,* Mahwah, NJ: Lawrence Erlbaum Associates.

Stemler, S. E., & Sternberg, R. J. (2006). Using situational judgment tests to measure practical intelligence. In J. A. Weekley & R. E. Ployhart (Eds.), *Situational judgment tests:*

Theory, measurement, and application (pp. 107-131). Mahwah, NJ: Lawrence Erlbaum Associates.

Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence*. New York, NY: Cambridge University Press.

Waugh, G. W., & Russell, T. L. (2005, April). Comparison of situational judgment tests formats, scoring key developers, and scoring algorithms. In F. L. Oswald (chair), *Understanding and improving situational judgment tests*. Symposium conducted at the meeting of the Society for Industrial and Organizational Psychology, Los Angeles, California.

Weekley, J. A., & Jones, C. (1997). Video-based situational testing. *Personnel Psychology, 50*, 25-49.

Weekley, J. A., & Jones, C. (1999). Further studies of situational tests. *Personnel Psychology, 52*, 679-700.

Weekley, J. A., & Ployhart, R. E. (2006). An introduction to situational judgment testing. In J. A. Weekley & R. E. Ployhart (Eds.), *Situational judgment tests: Theory, measurement, and application* (pp. 1-10). Mahwah, NJ: Lawrence Erlbaum Associates.

Weekley, J. A., Ployhart, R. E., & Holtz, B. C. (2006). On the development of situational judgment tests: Issues in item development, scaling, and scoring. In J. A. Weekley & R. E. Ployhart (Eds.), *Situational judgment tests: Theory, measurement, and application* (pp. 157-182). Mahwah, NJ: Lawrence Erlbaum Associates.

Woehr, D. J., & Arthur, W., Jr. (2003). The construct-related validity of assessment center ratings: A review and meta-analysis of the role of methodological factors. *Journal of Management, 29,* 231-258.

Wernimont, P. & Campbell, J. P. (1968). Signs, samples, and criteria. *Journal of Applied Psychology, 52*, 372-376.

APPENDIX

Table 1

	M	*SD*	1	2	3	4	5	6	7	8	9	10	11	12
1. Sex	1.40	.49	--											
2. Tenure	38.26	3.28	-.05	*.43*										
3. Education	26.26	3.59	.08	.00	*.74*									
4. Achievement	4.44	.46	.04	-.02	-.04	*.83*								
5. Adaptability	3.96	.55	.13	-.01	-.11	.55**	*.82*							
6. Analytical	4.02	.61	-.06	.06	-.28**	.64**	.51**	*.87*						
7. Concern	1.78	.52	-.14	-.17*	-.01	-.25**	-.25**	-.14	*.69*					
8. Cooperation	4.39	.44	.18*	.05	.10	.45**	.45**	.34**	-.61**	*.82*				
9. Dependability	4.41	.48	-.01	.07	.07	.56**	.37**	.41**	-.29**	.51**	*.82*			
10. Detail	3.99	.58	.10	-.03	-.01	.53**	.28**	.45**	-.03	.25**	.61**	*.81*		
11. Energy	4.11	.54	.06	-.01	-.09	.69**	.55**	.54**	-.11	.33**	.38**	.41**	*.83*	
12. Independence	2.99	.59	.02	.02	-.06	.07	-.15*	.13	.30**	-.20*	-.06	.21**	.09	*.71*
13. Initiative	4.12	.52	.06	.05	.02	.70**	.53**	.56**	-.14	.41**	.70**	.61**	.63**	.09
14. Innovation	3.92	.61	.09	-.10	-.07	.66**	.58**	.67**	-.09	.27**	.37**	.39**	.55**	.15
15. Integrity	2.13	.57	.04	-.08	-.05	-.33**	-.34**	-.17*	.35**	-.43**	-.43**	-.23**	-.24**	.18*
16. Leadership	3.79	.73	-.06	-.11	-.11	.57**	.49**	.58**	-.11	.25**	.33**	.33**	.58**	.04
17. Persistence	4.10	.49	.02	.08	-.10	.75**	.55**	.61**	-.16*	.38**	.63**	.59**	.68**	.08
18. Self-Control	2.14	.60	-.03	.04	.03	-.38**	-.49**	-.31**	.11	-.37**	-.36**	-.31**	-.32**	.17*
19. Social	4.05	.49	.04	.01	.09	.46**	.43**	.35**	-.54**	.68**	.39**	.14	.35**	-.36**
20. Stress	3.88	.61	.01	-.01	-.06	.40**	.58**	.44**	-.09	.36**	.27**	.27**	.52**	-.04
21. Performance	108.47	19.19	-.01	-.03	.05	-.08	.10	.11	.06	.01	-.02	.01	-.02	.01
22. Overall SJT	51.78	9.24	.10	.19*	.06	.12	.11	.12	-.13	.13	.13	.01	.04	.01
23. Leadership	9.47	2.95	.04	.14	-.10	.05	.08	.18*	-.00	-.04	.10	-.01	.15	.05
24. Interpersonal Skills	32.44	6.29	.18*	.12	-.02	.15	.11	.12	-.12	.10	.14	.06	.05	-.09
25. Sales Knowledge	9.87	2.99	.04	.15*	.15	.09	.07	.03	-.13	.17*	.09	-.02	-.04	.03

Table 1 continued

	13	14	15	16	17	18	19	20	21	22	23	24	25
13. Initiative	*.86*												
14. Innovation	.65**	*.91*											
15. Integrity	-.33**	-.16*	*.68*										
16. Leadership	.49**	.65**	-.18*	*.90*									
17. Persistence	.77**	.65**	-.37**	.55**	*.78*								
18. Self-Control	-.41**	-.31**	.45**	-.21**	-.45**	*.81*							
19. Social	.39**	.30**	-.43**	.35**	.42**	-.35**	*.76*						
20. Stress	.46**	.37**	-.34**	.35**	.48**	-.71**	.36**	*.82*					
21. Performance	-.01	.01	-.01	.05	.06	.02	-.04	-.02	--				
22. Overall SJT	.01	.08	-.08	.16*	.04	.02	.04	.02	.15	*.60*			
23. Leadership	.08	.07	-.01	.10	.10	-.01	-.10	.13	.02	.62**	*.10*		
24. Interpersonal Skills	.05	.11	-.04	.15	.12	-.01	.05	-.03	.01	.61**	.22**	*.21*	
25. Sales Knowledge	-.04	.03	-.09	.12	-.04	.04	.09	-.01	.06	.89**	.34**	.31**	*.50*

Note. For sex, 1= male and 2 = female. For all regressions, * p<.05, ** p<.01. Coefficient alphas are reported in the diagonal. Overall SJT: The best-and-worst scored SJT from Company 1. Concern: Concern for others, Detail: Detail orientation, Social: Social orientation, Stress: Stress tolerance. Leadership: The leadership subscale of the SJT. Interpersonal Skills: The interpersonal skills subscale of the SJT. Sales Knowledge: The sales knowledge subscale of the SJT. The sex, education, and performance variables were a single item measure, thus there is no alpha levels.

Table 2

	M	*SD*	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Sex	.45	.50	--												
2. Tenure	2.57	.73	-.03	*.53*											
3. Education	25.74	2.79	.03	-.06	.75										
4. Achievement	4.65	.36	.03	-.05	-.05	*.87*									
5. Adaptability	4.20	.47	-.01	-.10	.10	.49**	*.75*								
6. Analytical	4.31	.56	-.08	.04	-.07	.47**	.49**	*.90*							
7. Concern	1.68	.52	-.25**	.08	-.10	-.28**	-.43**	-.26**	*.72*						
8. Cooperation	4.47	.43	.26**	-.05	.08	.39**	.48**	.29**	-.73**	*.82*					
9. Dependability	4.53	.44	.02	.01	-.03	.56**	.50**	.43**	-.28**	.39**	*.81*				
10. Detail	4.20	.54	.08	-.00	-.06	.61**	.43**	.41**	-.20**	.36**	.60**	*.82*			
11. Energy	4.33	.53	-.05	-.08	.21**	.59**	.53**	.40**	-.30**	.42**	.57**	.57**	*.83*		
12. Independence	2.91	.56	-.07	.05	.04	-.03	-.29**	.09	.32**	-.31**	-.13	-.10	-.09	*.74*	
13. Initiative	4.27	.51	-.02	-.07	-.05	.66**	.56**	.56**	-.35**	.45**	.73**	.61**	.63**	-.09	*.87*
14. Innovation	4.18	.54	.02	-.04	.12	.48**	.52**	.64**	-.44**	.46**	.36**	.36**	.50**	-.01	.59**
15. Integrity	1.95	.59	.03	-.04	-.03	-.35**	-.32**	-.21**	.36**	-.44**	-.37**	-.40**	-.37**	.29**	-.39**
16. Leadership	4.23	.59	-.05	.10	-.01	.49**	.46**	.59**	-.31**	.27**	.44**	.43**	.51**	-.01	.52**
17. Persistence	4.29	.50	-.08	-.07	.03	.69**	.58**	.62**	-.27**	.33**	.65**	.57**	.62**	-.06	.74**
18. Self-Control	2.01	.64	.04	.10	-.11	-.36**	-.52**	-.36**	.28**	-.44**	-.42**	-.44**	-.35**	.22**	-.41**
19. Social	4.21	.43	.11	-.07	.06	.31**	.41**	.21**	-.52**	.64**	.34*	.29**	.37**	-.50**	.34**
20. Stress	4.10	.62	-.19*	-.08	.10	.41**	.67**	.53**	-.21**	.33**	.44**	.45**	.51**	-.20**	.52**
21. Performance	31.60	12.15	-.01	.08	-.08	.18*	.04	.15*	-.07	.02	.21**	.06	.07	-.00	.16*
22. Likert SJT	-230.33	34.34	-.05	-.06	-.09	-.12	-.17*	-.13	.24**	-.27**	-.24**	-.17	-.18*	.10	-.21**
23. Leadership	54.44	9.32	.11	-.03	.12	.24**	.29**	.19*	-.29**	.33**	.33**	.28**	.33**	-.11	.34**
24. Interpersonal Skills	50.74	8.43	.00	.08	.05	.12	.20**	.14	-.26**	.25**	.24**	.18*	.22**	-.10	.21**

Table 2 continued

	M	*SD*	1	2	3	4	5	6	7	8	9	10	11	12	13
25. Sales Knowledge	125.15	21.09	.04	.08	.07	.05	.06	.08	-.16*	.20**	.16*	.08	.06	-.06	.10
26. Converted SJT	53.84	8.55	.04	-.05	-.00	.01	.10	.06	-.01	-.05	.05	.02	.04	-.04	.08
27. Leadership Converted	11.68	3.31	-.10	-.01	-.10	-.10	.05	.02	.08	-.05	-.08	-.09	-.10	.04	-.04
28. Interpersonal Skills Converted	9.23	3.02	.25**	.05	-.03	.07	.03	.01	-.07	.07	.18*	.10	-.02	-.13	.14
29. Sales Knowledge Converted	32.93	5.93	-.01	-.09	.07	.03	.11	.07	-.02	-.08	.02	.03	.13	-.01	.06

Table 2 continued

	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
14. Innovation	*.89*															
15. Integrity	-.21**	*.76*														
16. Leadership	.57**	-.19*	*.88*													
17. Persistence	.56**	-.36**	.54**	*.84*												
18. Self-Control	-.32**	.38**	-.26**	-.53**	*.84*											
19. Social	.31**	-.37**	.34**	.29**	-.40**	*.71*										
20. Stress	.40**	-.36**	.39**	.64**	-.71**	.38**	*.85*									
21. Performance	.07	.04	.11	.21**	-.08	-.03	.02	--								
22. Likert SJT	-.22**	.16*	-.20*	-.27**	.08	-.14	-.09	.05	*.88*							
23. Leadership	.24**	-.26**	.13	.32**	-.13	.17*	.14	.08	.76**	*.68*						
24. Interpersonal Skills	.21**	-.20**	.05	.33**	-.11	.13	.14	.05	.86	.62**	*.63*					
25. Sales Knowledge	.16*	-.06	-.04	.17*	-.02	.10	.02	.01	.94**	.55**	.71**	*.81*				
26. Converted SJT	.04	-.06	.22**	.01	-.05	.03	.04	-.04	.51**	-.35**	-.34**	-.54**	*.47*			
27. Leadership Converted	-.01	.05	.08	-.03	-.04	-.07	.00	.36**	-.13	-.36**	.05	.35**	.63**	*.43*		
28. Interpersonal Skills Converted	.05	.00	.14	.01	.03	.18*	-.04	.21**	-.31**	-.16*	.13	.23**	.46**	.08	*.37*	
29. Sales Knowledge Converted	.04	-.12	.21**	.03	-.07	-.01	.08	.21**	-.26**	-.50**	-.04	.43**	.85**	.31**	.11	*.58*

Note. For sex, 1 = male and 2 = female. For all regressions, * p<.05, ** p<.01. Coefficient alphas are reported in the diagonal. Likert SJT: The Likert scored SJT from Company 2, Converted SJT: The Likert scored SJT that was converted to a best-and-worst scored SJT from Company 2. Concern: Concern for others, Detail: Detail orientation, Social: Social orientation, Stress: Stress tolerance. Leadership and Leadership Converted: The leadership subscale of the SJT. Interpersonal Skills and Interpersonal Skills Converted: The interpersonal skills subscale of the SJT. Sales Knowledge and Sales Knowledge Converted: The sales knowledge subscale of the SJT. The sex, education, and performance variables were a single item measure, thus there is no alpha levels.

Table 3

		Best/Worst SJT					Likert SJT					Converted SJT				
Overall SJT																
Step		*β*	*R²*	*F*	*ΔR²*	*ΔF*	*β*	*R²*	*F*	*ΔR²*	*ΔF*	*β*	*R²*	*F*	*ΔR²*	*ΔF*
1	Sex	-.11	.03	2.07	.03	2.07	.08	.01	.60	.01	.60	.08	.01	.60	.01	.60
	Tenure	-.12					-.00					-.00				
2	SJT	.20*	.06	3.54*	.04	6.34*	.04	.01	.51	.00	.34	.06	.02	.63	.01	.88
Leadership																
Step		*β*	*R²*	*F*	*ΔR²*	*ΔF*	*β*	*R²*	*F*	*ΔR²*	*ΔF*	*β*	*R²*	*F*	*ΔR²*	*ΔF*
1	Sex	-.01	.00	.02	.00	.02	.08	.01	.60	.01	.60	.08	.01	.60	.01	.60
	Tenure	.01					-.00					-.00				
2	SJT	.03	.01	.06	.00	.14	.07	.02	.72	.01	.97	.07	.01	.65	.00	.75
Interpersonal Skills																
Step		*β*	*R²*	*F*	*ΔR²*	*ΔF*	*β*	*R²*	*F*	*ΔR²*	*ΔF*	*β*	*R²*	*F*	*ΔR²*	*ΔF*
1	Sex	-.01	.00	.02	.00	.02	.08	.01	.60	.01	.60	.08	.01	.60	.01	.60
	Tenure	.01					-.00					-.00				
2	SJT	.03	.01	.04	.00	.09	-.05	.01	.55	.00	.46	.14	.03	1.57	.02	3.49
Sales Knowledge																
Step		*β*	*R²*	*F*	*ΔR²*	*ΔF*	*β*	*R²*	*F*	*ΔR²*	*ΔF*	*β*	*R²*	*F*	*ΔR²*	*ΔF*
1	Sex	-.01	.00	.02	.00	.02	.08	.01	.60	.01	.60	.08	.01	.60	.01	.60
	Tenure	.01					-.00					-.00				
2	SJT	.07	.01	.27	.01	.77	-.02	.01	.42	.00	.06	-.01	.01	.41	.00	.04

Note. For all regressions, * p<.05. Best/Worst SJT: The best-and-worst SJT from Company 1, Likert SJT: The Likert SJT from Company 2, Converted SJT: The Likert SJT that was converted to a best-and-worst scored SJT from Company 2. Leadership: The leadership subscale of the SJT. Interpersonal Skills: The interpersonal skills subscale of the SJT. Sales Knowledge: The sales knowledge subscale of the SJT. For the overall Best/Worst SJT and each of the subscale: *df*(step 1) = 2, 156, *df*(step 2) = 3, 155. For the overall Likert SJT and each of the subscales as well as the converted Likert SJT and the subscales: *df*(step 1) = 2, 175, *df*(step2) = 3, 174.

Table 4

		Best/Worst SJT					Likert SJT					Converted SJT				
Step		*β*	R^2	*F*	ΔR^2	ΔF	*β*	R^2	*F*	ΔR^2	ΔF	*β*	R^2	*F*	ΔR^2	ΔF
1	Sex	.11	.05	4.23*	.05	4.23*	-.06	.01	.56	.01	.56	-.05	.00	.29	.00	.29
	Tenure	.20*					-.05					.03				
2	Education	.04	.05	2.88*	.00	.23	-.09	.03	.80	.01	1.29	-.01	.00	.19	.00	.01
3	O*Net competencies		.18	1.60	.13	1.31		.25	2.38**	.22	2.59		.14	1.41	.14	1.54
	Achievement	.26*					.21					-.07				
	Adaptability	.08					-.02					.14				
	Analytical	.05					.03					-.03				
	Concern	.02					.10					.06				
	Cooperation	.11					-.16					-.16				
	Dependability	.28*					-.17					-.06				
	Detail	-.14					-.11					-.13				
	Energy	-.14					-.01					.01				
	Independence	.03					.00					-.04				
	Initiation	-.22					.12					.21				
	Innovation	-.08					-.15					-.09				
	Integrity	-.03					-.02					-.11				
	Leadership	.24*					.29**					.43**				
	Persistence	-.14					-.48**					-.23				
	Self-Control	.14					-.20					-.19				
	Social	-.16					-.07					-.07				
	Stress	.09					.09					-.14				

Note. For all regressions, * p<.05, ** p<.01. Best/Worst SJT: The best-and-worst scored SJT from Company 1, Likert SJT: The Likert scored SJT from Company 2, Converted SJT: The Likert scored SJT that was converted to a best-and-worst scored SJT from Company 2. Concern: Concern for others, Detail: Detail orientation, Social: Social orientation, Stress: Stress tolerance. Best/Worst SJT: *df*(step 1) = 2, 157, *df*(step 2) = 3, 156, *df*(step3) = 20, 139. Likert SJT: *df*(step 1) = 2, 175, *df*(step2) = 3, 174, *df*(step 3) = 19, 158. Converted SJT: *df*(step 1) = 2, 175, *df*(step2) = 3, 174, *df*(step3) = 19, 158.

Table 5

		Best/Worst SJT					Likert SJT					Converted SJT				
Step		*β*	R^2	*F*	ΔR^2	ΔF	*β*	R^2	*F*	ΔR^2	ΔF	*β*	R^2	*F*	ΔR^2	ΔF
1	Sex	.05	.02	1.69	.02	1.69	-.03	.01	.94	.01	.94	-.01	.01	.89	.01	.89
	Biodata	.14					.10					-.10				
2	Education	-.12	.03	1.90	.01	2.29	.13	.03	1.56	.02	2.77	-.09	.02	1.08	.01	1.45
3	O*Net competencies		.18	1.57	.15	1.54		.26	2.61**	.23	2.79		.12	1.02	.10	1.03
	Achievement	-.08					-.15					-.15				
	Adaptability	-.01					.12					.23				
	Analytical	.31*					-.03					-.01				
	Concern	-.06					-.10					.16				
	Cooperation	-.14					.14					.14				
	Dependability	.27*					.07					-.12				
	Detail	-.26*					.10					-.08				
	Energy	.12					.13					-.16				
	Independence	-.01					-.02					.01				
	Initiation	.03					.08					.13				
	Innovation	-.15					.02					-.09				
	Integrity	-.01					-.10					.01				
	Leadership	.06					-.16					.27*				
	Persistence	-.00					.24					.02				
	Self-Control	.08					.16					-.20				
	Social	-.26*					-.06					.10				
	Stress	.16					-.07					-.19				

Note. For all regressions, * p≤.05, ** p<.01. Best/Worst SJT: The best-and-worst scored SJT from Company 1, Likert SJT: The Likert scored SJT from Company 2, Converted SJT: The Likert scored SJT that was converted to a best-and-worst scored SJT from Company 2. Concern: Concern for others, Detail: Detail orientation, Social: Social orientation, Stress: Stress tolerance. Best/Worst SJT: *df*(step 1) = 2, 157, *df*(step 2) = 3, 156, *df*(step3) = 19, 140. Likert SJT: *df*(step 1) = 2, 175, *df*(step2) = 3, 174, *df*(step 3) = 19, 158. Converted SJT: *df*(step 1) = 2, 175, *df*(step2) = 3, 174, *df*(step3) = 19, 158.

Table 6

		Best/Worst SJT					Likert SJT					Converted SJT				
Step		*β*	*R²*	*F*	*ΔR²*	*ΔF*	*β*	*R²*	*F*	*ΔR²*	*ΔF*	*β*	*R²*	*F*	*ΔR²*	*ΔF*
1	Sex	.19	.05	4.06*	.05	4.06	.08	.01	.57	.01	.57	.06	.06	5.83**	.06	5.83
	Biodata	.13					.00					.24				
2	Education	-.06	.05	2.88*	.00	.56	.04	.01	.49	.00	.31	-.03	.06	3.94**	.00	.22
3	O*Net competencies		.15	1.30	.10	.97		.27	2.95**	.26	3.21		.17	1.76*	.11	1.26
	Achievement	.21					-.28*					-.01				
	Adaptability	.02					.09					-.03				
	Analytical	.15					-.05					-.07				
	Concern	.02					-.17					.09				
	Cooperation	.02					.08					-.08				
	Dependability	.16					.09					.19				
	Detail	-.06					.13					.03				
	Energy	-.17					.05					-.21				
	Independence	-.14					.02					-.08				
	Initiation	-.19					-.13					.19				
	Innovation	-.07					.09					.05				
	Integrity	.02					-.01					.05				
	Leadership	.17					-.27**					.16				
	Persistence	.08					.63**					-.11				
	Self-Control	.02					.17					.08				
	Social	-.13					.05					.17				
	Stress	-.07					-.12					.00				

Note. For all regressions, * p<.05, ** p<.01. Best/Worst SJT: The best-and-worst scored SJT from Company 1, Likert SJT: The Likert scored SJT from Company 2, Converted SJT: The Likert scored SJT that was converted to a best-and-worst scored SJT from Company 2. Concern: Concern for others, Detail: Detail orientation, Social: Social orientation, Stress: Stress tolerance. Best/Worst SJT: *df*(step 1) = 2, 157, *df*(step 2) = 3, 156, *df*(step3) = 19, 140. Likert SJT: *df*(step 1) = 2, 175, *df*(step2) = 3, 174, *df*(step 3) = 19, 158. Converted SJT: *df*(step 1) = 2, 175, *df*(step2) = 3, 174, *df*(step3) = 20, 157.

Table 7

		Best/Worst SJT					Likert SJT					Converted SJT				
Step		*β*	R^2	*F*	ΔR^2	ΔF	*β*	R^2	*F*	ΔR^2	ΔF	*β*	R^2	*F*	ΔR^2	ΔF
1	Sex	.05	.03	2.48	.03	2.48	.08	.01	.69	.01	.69	-.09	.01	.72	.01	.72
	Biodata	.17					.04					-.02				
2	Education	.14	.05	2.72*	.02	3.14	.07	.02	.73	.01	.81	.06	.01	.68	.00	.59
3	O*Net competencies		.21	1.68*	.16	1.57		.19	1.76*	.17	1.88		.18	1.85*	.17	1.97
	Achievement	.31*					-.16					-.01				
	Adaptability	.12					-.05					.09				
	Analytical	-.14					-.02					-.00				
	Concern	.05					-.06					-.05				
	Cooperation	.22					.17					-.27*				
	Dependability	.20					.20					-.11				
	Detail	-.05					.09					-.15				
	Energy	-.18					-.06					.20				
	Independence	.11					.00					-.03				
	Initiation	-.25					-.17					.14				
	Innovation	-.01					.21					-.10				
	Integrity	-.05					.07					-.20*				
	Leadership	.24*					-.30*					.39**				
	Persistence	-.24					.42*					-.29*				
	Self-Control	.15					.19					-.21				
	Social	-.05					.12					-.13				
	Stress	.08					-.08					-.09				

Note. For all regressions, * p<.05, ** p<.01. Best/Worst SJT: The best-and-worst scored SJT from Company 1, Likert SJT: The Likert scored SJT from Company 2, Converted SJT: The Likert scored SJT that was converted to a best-and-worst scored SJT from Company 2. Concern: Concern for others, Detail: Detail orientation, Social: Social orientation, Stress: Stress tolerance. Best/Worst SJT: *df*(step 1) = 2, 157, *df*(step 2) = 3, 156, *df*(step3) = 19, 140. Likert SJT: *df*(step 1) = 2, 175, *df*(step2) = 3, 174, *df*(step 3) = 19, 158. Converted SJT: *df*(step 1) = 2, 175, *df*(step2) = 3, 174, *df*(step3) = 19, 158.

Table 8

		Best/Worst SJT					Likert SJT					Converted SJT				
Step		*β*	R^2	*F*	ΔR^2	ΔF	*β*	R^2	*F*	ΔR^2	ΔF	*β*	R^2	*F*	ΔR^2	ΔF
1	Sex	-.01	.03	.01	.03	.02	-.00	.01	.60	.01	.60	-.00	.01	.60	.01	.60
	Biodata	.01					.08					.08				
2	Education	.06	.03	.21	.00	.59	-.07	.01	.66	.00	.78	-.07	.01	.66	.00	.78
3	O*Net competencies		.14	.73	.11	.82		.15	1.35	.14	1.47		.15	1.35	.14	1.47
	Achievement	-.33*					.15					.15				
	Adaptability	.25					-.12					-.12				
	Analytical	.25					.14					.14				
	Concern	.07					-.16					-.16				
	Cooperation	.08					-.10					-.10				
	Dependability	-.06					.22					.22				
	Detail	.04					-.13					-.13				
	Energy	.09					.01					.01				
	Independence	.04					-.11					-.11				
	Initiation	.05					.03					.03				
	Innovation	.02					-.09					-.09				
	Integrity	.03					.14					.14				
	Leadership	-.06					-.06					-.06				
	Persistence	-.10					.17					.17				
	Self-Control	-.09					-.13					-.13				
	Social	.05					-.09					-.09				
	Stress	-.23					-.13					-.13				
4	SJT	.08	.18	.73	.04	.78	.14	.17	1.45	.02	3.01	.09	.16	1.35	.01	1.23
4	Leadership	.04	.14	.70	.00	.21	-.17	.17	1.52	.02	4.20	.10	.16	1.36	.01	1.49
4	Interpersonal	.04	.14	.70	.00	.17	-.17	.17	1.51	.02	4.11	.13	.17	1.44	.02	2.76
4	Sales	.08	.15	.73	.01	.78	-.09	.16	1.35	.01	1.24	.94	.15	1.28	.00	.01

Note. For all regressions, * p<.05. Best/Worst SJT: The best-and-worst scored SJT from Company 1, Likert SJT: The Likert scored SJT from Company 2, Converted SJT: The Likert scored SJT that was converted to a best-and-worst scored SJT from Company 2. Concern: Concern for others, Detail: Detail orientation, Social: Social orientation, Stress: Stress tolerance. Best/Worst SJT: *df*(step 1) = 2, 156, *df*(step 2) = 3, 155, *df*(step 3) = 20, 138, *df*(step 4) = 21, 137. Likert SJT: *df*(step 1) = 2, 175, *df*(step 2) = 3, 174, *df*(step 3) = 20, 157, *df*(step 4) = 21, 156. Converted SJT: *df*(step 1) = 2, 175, *df*(step 2) = 3, 174, *df*(step 3) = 20, 157, *df*(step 4) = 21, 156.

VITA

Name: Jennifer Lynn Rasmussen

Address: Texas A&M University
Department of Psychology
College Station, TX 77843-4235
(979) 845-0487

E-mail: jlr3723@tamu.edu

Education:

2009	M.S. in Psychology, Texas A&M University
2005	B.S. in Psychology with Honors, Business Minor Florida State University (2005) Magna cum laude, Phi Beta Kappa Thesis Title: Reducing Backlash with Social Approval: The Implication of Narcissism Certificate in Performance Management